View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Current Biology 20, 895–903, May 25, 2010 ©2010 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2010.04.027

Article

Rapid Expansion and Functional Divergence of Subtelomeric Gene Families in Yeasts

Chris A. Brown,[1] Andrew W. Murray,[1] and Kevin J. Verstrepen[1,2,3,*]

[1]Faculty of Arts and Sciences Center for Systems Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA

[2]VIB Laboratory for Systems Biology

[3]Centre of Microbial and Plant Genetics (CMPG-G&G), K.U.Leuven

Gaston Geenslaan 1, B-3001 Leuven (Heverlee), Belgium

Summary

Background: Subtelomeres, regions proximal to telomeres, exhibit characteristics unique to eukaryotic genomes. Genes residing in these loci are subject to epigenetic regulation and elevated rates of both meiotic and mitotic recombination. However, most genome sequences do not contain assembled subtelomeric sequences, and, as a result, subtelomeres are often overlooked in comparative genomics.

Results: We studied the evolution and functional divergence of subtelomeric gene families in the yeast lineage. Our computational results show that subtelomeric families are evolving and expanding much faster than families that do not contain subtelomeric genes. Focusing on three related subtelomeric *MAL* gene families involved in disaccharide metabolism that show typical patterns of rapid expansion and evolution, we show experimentally how frequent duplication events followed by functional divergence yield novel alleles that allow the metabolism of different carbohydrates.

Conclusions: Taken together, our computational and experimental analyses show that the extraordinary instability of eukaryotic subtelomeres supports rapid adaptation to novel niches by promoting gene recombination and duplication followed by functional divergence of the alleles.

Introduction

Subtelomeres are repeat-rich and gene-poor regions proximal to the telomeres [1]. A precise definition of a subtelomere is difficult because the length of the subtelomeric region varies from 20 kb in some yeasts to several hundred kb in higher eukaryotes [2, 3]. Apart from the low gene density, subtelomeres are characterized by epigenetic silencing [4, 5] and increased rates of recombination and mutation [3, 6–9], with the exception of flies [10, 11]. These regions are often lacking from so-called whole-genome sequences because their high repeat content and extensive sequence similarity [12] make it difficult to assemble these regions and to distinguish orthologs and paralogs [2, 13, 14]. As a result, subtelomeres remain relatively understudied. For example, several landmark studies that reconstruct the evolution of gene families could not comprehensively analyze subtelomeric gene families [14–17]. From the few examples we have, subtelomeres seem to contain specific gene families that reflect the organism's lifestyle. In yeasts, genes involved in biofilm formation and carbohydrate utilization have been mapped to subtelomeres [18–23]. In parasitic eukaryotes such as *Plasmodium spp*., trypanosomes, and pathogenic fungi, many virulence genes reside at subtelomeres. Variegated expression of these cell-surface genes allows these pathogens to continuously change their outer surface and evade the host immune response [24, 25]. In primates, multiple genes encoding olfactory receptors [26] and members of the WASP family [27] have been mapped to subtelomeres. Moreover, promiscuous rearrangements of these regions have been implicated in human genetic disorders [28, 29]. These anecdotal examples support the hypothesis that subtelomeres are variable loci harboring specific and fast-evolving gene families. Indeed, other authors have noted the rapid turnover of genes at subtelomeres [30–32], but a comprehensive analysis is lacking. Here, we use the genome sequences of eight ascomycete fungi to study the evolution of subtelomeric genes. We use comparative genomics to show that subtelomeric gene families evolve faster than their nonsubtelomeric counterparts and then focus on three related gene families to analyze how they have evolved and functionally diverged. Together, our results underpin the unique role of subtelomeres as hotbeds for genomic evolution and innovation.

Results

In Silico Analysis of Subtelomeric Gene Families

Using the definition of subtelomeres as gene-depleted regions [33], we investigated the gene density across the chromosomes of various yeast species and found that the average gene density is significantly lower up to 33 kb away from the telomeres (see Figure S1 available online). This 33 kb region agrees with previous studies about the telomere position effect [5] and sequence similarity between nonhomologous chromosome ends [2, 3]. For our analyses, we classified each gene as subtelomeric or nonsubtelomeric based on its distance from the chromosome end. Although we only show results for a telomere length of 33 kb as defined above, all of our results are robust for subtelomere lengths between 10 and 50 kb (see below).

To investigate which genes are enriched and depleted at subtelomeres, we used the gene ontology (GO) classification [34] (Table S1). Subtelomeres show significant enrichment for genes involved in response to stress and toxins, metabolism of a broad spectrum of compounds, and transporters involved in metal, amino acid, and carbohydrate uptake. By contrast, genes responsible for typical housekeeping functions such as ribosomal function, RNA processing, cell-cycle control, mitosis, DNA repair, and DNA replication are depleted at subtelomeres.

In order to compare the evolution of subtelomeric and nonsubtelomeric genes, we used the Markov cluster algorithm (MCL algorithm) [15, 35, 36] to divide all genes across different fungi into gene families based on their sequence similarity (see Experimental Procedures). Gene families that contain at least one gene located in the subtelomeric region were considered to be subtelomeric families. On average, gene families that do

*Correspondence: kevin.verstrepen@biw.vib-kuleuven.be

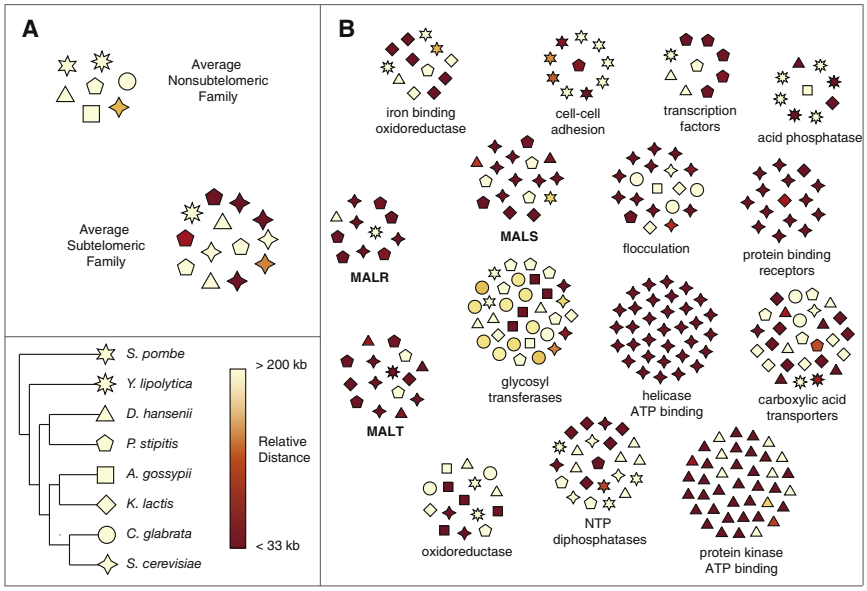

Figure 1. Overview of Subtelomeric Families versus Nonsubtelomeric Families

Subtelomeric families are more volatile than nonsubtelomeric families. Every gene family (orthogroup) is shown as a number of polygons (individual genes per gene family) spatially arranged in a circular area. The polygons are colored according to their closest distance (denoted as relative distance in the figure) to a chromosome end going from pale yellow (>200 kb away from the nearest chromosome end) to dark red (<33 kb from the nearest chromosome end). The individual species are denoted by the eight differently shaped polygons (e.g., triangle for *D. hansenii*).

(A) The average composition of nonsubtelomeric gene families and subtelomeric gene families is represented by two clusters (artificial clusters representing the mean copy number and mean distance to the telomere for subtelomeric and nonsubtelomeric gene families). Nonsubtelomeric gene families show small differences in copy number between species such that species contain around the same number of genes and show few genes within 200 kb of the chromosome end. Strikingly, subtelomeric gene families show high copy number variation between species and have more genes within 200 kb (and especially within 33 kb) of the chromosome end.

(B) Some representative subtelomeric gene families are shown along with their functional annotation. Common characteristics among subtelomeric gene families can be seen: high copy number variation can be seen between species, as well as multiple members less than 200 kb from the nearest chromosome end. The three *MAL* gene families, *MALR*, *MALT*, and *MALS*, are shown in bold text. See Figure S1 and Table S1 for more information.

not contain any subtelomeric genes show a small number of genes per species, with only few genes located within 200 kb from the chromosome ends and very little difference in copy number between species. Subtelomeric gene families (i.e., families that contain at least one subtelomeric gene), on the other hand, often show several genes within 33 kb of the chromosome end and even more genes within 200 kb. Moreover, subtelomeric families also show drastic copy number variation between the different yeast species used in this analysis (Figure 1A).

Statistical analysis of the gene families shows that within a species, there are far fewer subtelomeric gene families than would be expected if subtelomeric and nonsubtelomeric genes were distributed randomly among the families. In other words, subtelomeric genes cluster together in a small number of families, and families that contain at least one subtelomeric member are more likely to contain multiple subtelomeric members ($p < 10^{-10}$) (Figure 1B; Figure 2A; Table S2); this signal remains even after controlling for tandem or local duplications ($p < 10^{-10}$) (Table S2). Even more striking, subtelomeric gene families are on average much larger than nonsubtelomeric families, containing 2–4 times more genes than nonsubtelomeric families ($p < 10^{-10}$)(Figure 1B; Figure 2B; Table S2).

Together, these analyses suggest that subtelomeric genes tend to spawn new subtelomeric genes, possibly as a result of the elevated recombination frequencies found at subtelomeres [3, 6, 7, 37]. This hypothesis prompted us to ask whether subtelomeric families also show more copy number variation than their nonsubtelomeric counterparts. When comparing gene family size across the fungal tree, subtelomeric families show significantly greater copy number variation between species ($p < 10^{-10}$) (Figure 1B; Figure 2C). Moreover, subtelomeric families contain many genes that show greater similarity to other genes in the same species than to genes in all other species: a signature for recent duplication events that occurred after the different species diverged ($p < 10^{-10}$) (Figure 2D). The extraordinarily rapid evolution of subtelomeric gene families was further investigated via the computational analysis of gene family evolution (CAFE) birth/death model [15]. This model uses a clever algorithm to quantify the rate at which new members of a gene family are being formed or lost (see Supplemental Experimental Procedures for more information). The model confirms that subtelomeric families show remarkably aberrant birth/death rates ($p < 10^{-10}$) (Table S2; Figure S2), further demonstrating the rapid evolution of these families.

After demonstrating that subtelomeric gene families show both elevated copy number variation and gene family size, we revisited our initial GO enrichment analysis. We wondered whether the rapid gene turnover at subtelomeres is a property of the types of genes found at subtelomeres or rather a property of the subtelomeric region. We therefore compared copy number variation and family size of nonsubtelomeric gene families to subtelomeric gene families belonging to the same functional GO category, and we repeated this analysis for all GO categories that are enriched for subtelomeric genes. We found that for almost all families of a specific category (99% and 98% of nonsubtelomeric and subtelomeric families tested, respectively), subtelomeric gene families showed both higher copy number variation and average family size. In some cases, we did not find statistically significant differences, but this seemed to be due to the low numbers of genes in these GO categories. Taken together, these results indicate that the subtelomeric location rather than the functional enrichment is the causal driving force for the rapid gene turnover, with frequent duplication events (Table S1).

Duplication of Subtelomeric Genes Involved in Disaccharide Utilization

Gene duplication is recognized as a crucial mechanism in evolution. The extra copy resulting from duplication events provides a dispensable copy of a gene that can acquire new function (neofunctionalization) without being restrained by

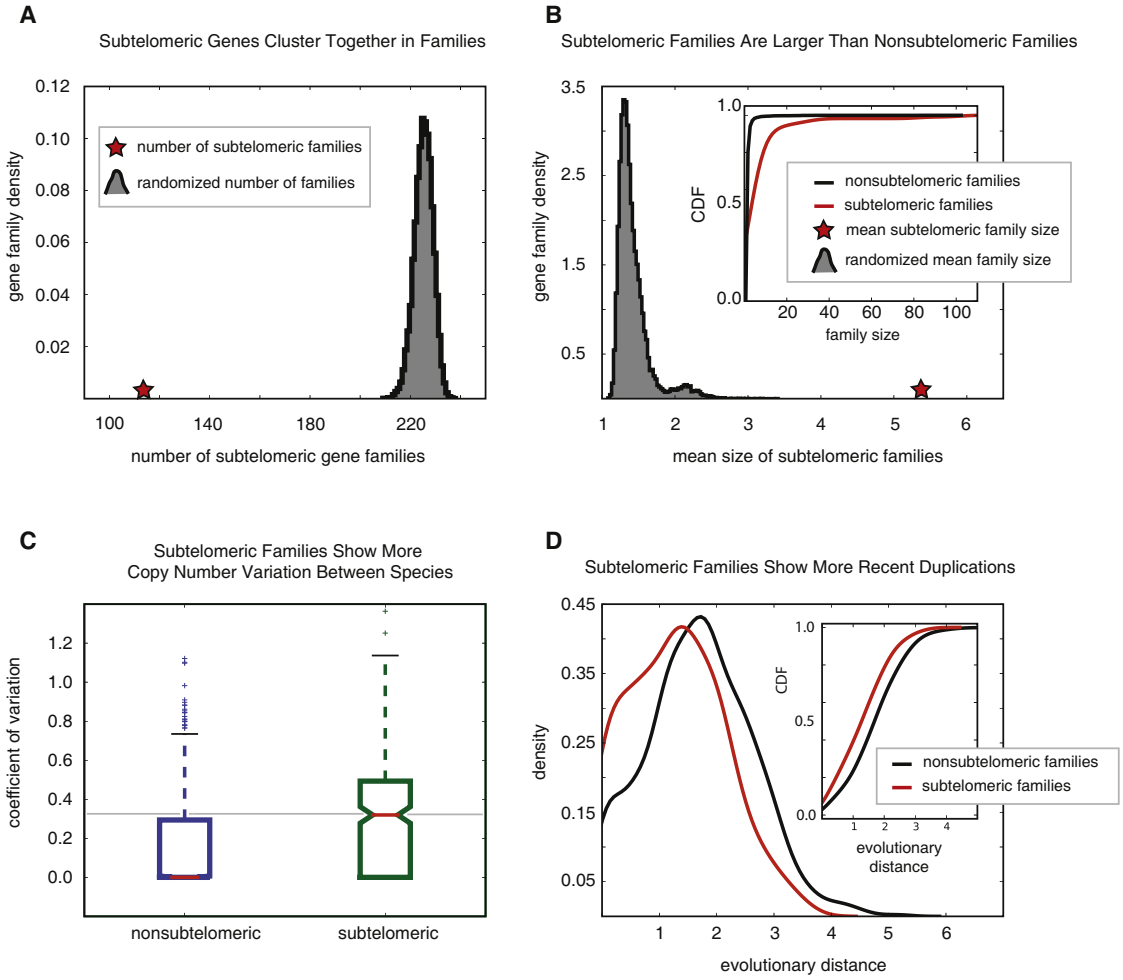

Figure 2. Computational Results

Subtelomeric gene families are larger and evolve faster. All *Saccharomyces cerevisiae* genes were divided into gene families based on their homology (see Results for details). Families containing at least one gene located within 33 kb of a telomere were classified as subtelomeric gene families; all other families are nonsubtelomeric families. Although we used the same 33 kb cutoff and the same Markov cluster (MCL) clustering parameters for our analyses, our results remained unchanged when we altered the definition of the subtelomeric region and also when we altered the parameters for MCL gene clustering.

(A) The total number of subtelomeric gene families (114, red star) is smaller than what would be expected if all genes (subtelomeric and nonsubtelomeric) were randomly distributed among gene families (gray curve representing the number of gene families with subtelomeric members after 10,000 randomizations).

(B) Subtelomeric gene families on average contain 2–4 times more genes than nonsubtelomeric families. The red star represents the average family size of subtelomeric families in *S. cerevisiae* at a 33 kb cutoff. The gray distribution shows the mean size of 10,000 gene families that were chosen randomly among all gene families. A complementary analysis contrasting the cumulative distribution functions of subtelomeric gene family size (red) and nonsubtelomeric gene family size (black) is consistent with the larger size of subtelomeric families.

(C) Subtelomeric gene families show increased copy number variation. The distribution of coefficients of variation (standard deviation normalized by mean, a dispersion metric) of the number of genes for all subtelomeric and nonsubtelomeric gene families shows that subtelomeric gene families exhibit drastically higher copy number variation than nonsubtelomeric gene families.

(D) The distribution of intraspecies protein distances (similarities) is compared between subtelomeric gene families (red) and nonsubtelomeric gene families (black). Subtelomeric gene families contain more closely related intraspecies proteins (recent duplications) than nonsubtelomeric gene families, which is reflected in a shift of the distribution to the left. The inset compares the cumulative distribution functions of the intraspecies protein distances and shows a significant shift of the subtelomeric families toward newer duplications. See Figure S2 and Table S2 for more information.

purifying selective pressure on its original function [38]. An alternative view is that gene duplication allows asymmetric evolution of preexisting promiscuous functions in a protein such that these prior functions can be further optimized (subfunctionalization) [39]. Another putative advantage of subfunctionalization is that the expression of the two copies can be independently regulated, which further increases the evolutionary potential.

To begin investigating whether members of subtelomeric gene families show signs of functional divergence, we studied their expression divergence (a measure for how differently the genes are regulated; see [40]) and responsiveness (a measure for how strongly a gene's expression is influenced by the environment [40]) (see Experimental Procedures). The results show that subtelomeric genes show higher average expression divergence (0.250 versus −0.007; $p = 0.035$) and higher average responsiveness (1961 versus 1491; $p < 10^{-10}$) when compared to nonsubtelomeric genes, agreeing with the hypothesis that subtelomeric duplicates show rapid divergence.

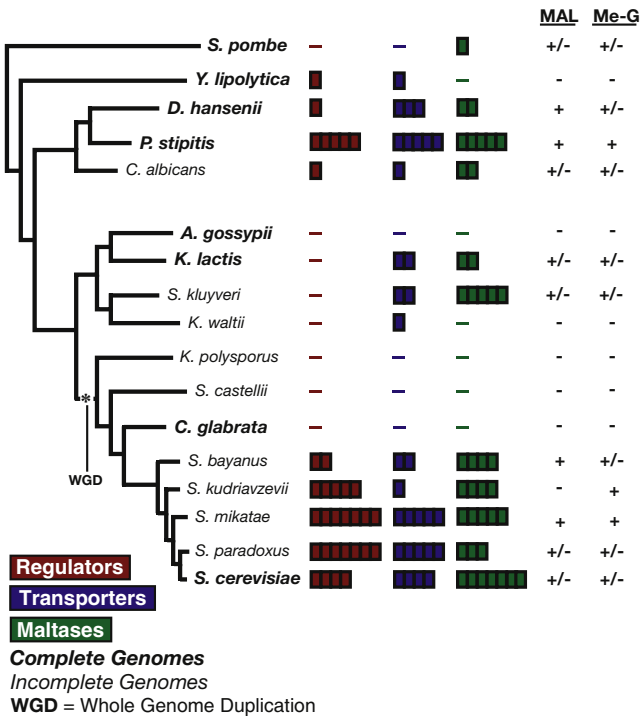

Figure 3. *MAL* Copy Number Variation in Fungal Lineage

The *MAL* gene family shows extreme copy number variation in yeasts, and the presence of *MAL* genes correlates with the ability to grow on maltose and methyl-α-glucoside (Me-G). The species names of genomes that have been completely sequenced and assembled are indicated in bold. The numbers of *MAL* regulator genes (red), *MAL* transporter genes (blue), and *MAL* maltase genes (green) are denoted by the number of blocks to the right of the species name. The panel on the right indicates whether (+) or not (−) individual strains grow on maltose (MAL) and methyl-α-glucoside. Strain-dependent growth is denoted as +/−. See Figure S3 and Table S3 for more information.

To further investigate whether the frequent duplication of subtelomeric genes provides the raw material for functional divergence, we examined three related typical subtelomeric gene families involved in maltose metabolism [41] (see Figure 1). For each of these families, we investigated whether and how the genes have been duplicated and whether these duplication events were followed by functional divergence. The first family, called *MALT*, contains transporters to import maltose into the cell; the second family, *MALS*, encodes maltases, enzymes that hydrolyze maltose into two glucose units; and the third family, *MALR*, encodes regulator proteins that induce the expression of *MALR*, *MALT*, and *MALS* genes when maltose is present [41].

We first manually mapped all *MAL* genes in completely assembled yeast genomes, as well as in available contigs of other (nonassembled) high-coverage genomes. We identified 7 unannotated *MAL* genes (2 from the *MALR* family and 5 from the *MALS* family) out of a total of 14 *MAL* genes in the *S. cerevisiae* S288c genome that were present as unannotated open reading frames. Next, consistent with our in silico analysis, we noted extraordinary fluctuations in the chromosomal location and number of *MAL* genes between different species and even strains (Figure 3; Figure 4; Figure S4). These copy number variations are not a direct result of the whole-genome duplication that occurred during the evolution of the hemiascomycetes [42]. *Candida glabrata*, *Saccharomyces castelli*, and *Kluyveromyces polysporus* underwent the whole-genome duplication but do not have any *MAL* loci. The protein phylogeny indicates that the common ancestor of these yeasts had only few *MAL* genes, which were completely lost in some lineages and expanded in other lineages (Figure S3).

Further phylogenetic analysis revealed the existence of multiple subfamilies (clades) of the *MALT*, *MALS*, and *MALR* families that cluster tightly together based on their sequence similarity (Figure 4; Figure S3). Genes within one subfamily do not only represent orthologs (i.e., copies that diverged independently after they were separated by speciation events; no gene duplication involved) but also represent recent paralogs (i.e., copies generated in duplication events within the species). Members of different subfamilies, on the other hand, show more sequence divergence and are usually ancient orthologs. Hence, the *MAL* genes show a remarkable instability in copy number and genomic location, even between evolutionary closely related *S. cerevisiae* strains. These characteristics of the different *MAL* genes agree very well with the results of our global in silico analysis of all subtelomeric genes (above). It is important to stress that we only based our analyses on the available fully sequenced yeast species. However, analysis of the *MAL* gene families in as many as 76 other (partly) sequenced *S. cerevisiae* and *S. paradoxus* strains confirm the trends observed in the fully assembled genomes (Table S3).

Functional Divergence in the *MAL* Gene Families

Given the rapid expansion of the *MAL* gene families in *S. cerevisiae*, we asked whether the duplication events resulted in sub- and/or neofunctionalization. We screened three sequenced *S. cerevisiae* strains for their ability to grow on maltose and other related carbohydrates. Our systematic analysis extended previous work [41, 43–46] and uncovered many novel functions for the different *MAL* genes. The laboratory strain S288c failed to grow on maltose, whereas two feral isolates, RM11 (from a vineyard) and YJM789 (from an AIDS patient), both grew. Further analysis showed that this difference depended on the absence of one specific *MALR* subfamily (clade) from S288c (Figure 4). Expressing members of the *MAL63*-like subfamily (*MAL63c9*, *MAL63c2* from RM11, and *MALx3* from YJM789) in S288c restored growth on maltose. Conversely, deleting all members of this subfamily in strains RM11 and YJM789 abolished their capacity to ferment maltose (Figure 5A). Further growth assays showed that these regulators are also required for growth on turanose, maltotriose, methyl-α-glucoside, isomaltose, palatinose, and sucrose (Figure 5B; Figure S5).

A phylogeny of the proteins encoded by the *MALR* genes shows three distinct subgroups of regulators. The previous results show that one *MALR* clade is vital for the consumption of many α-glucosides. Why, then, are the other regulator clades maintained? Further tests revealed that other regulators (*YFL052W*, *MAL13*, and *MAL33* in the *MAL13*-like clade) evolved specificity for palatinose, a disaccharide naturally occurring in sugarcane and honey (Figure 4; Figure S5C). Together, these results indicate that the functions of the different *MALR* paralogs have diverged to regulate cellular metabolism in response to various distinct carbohydrates.

Next, we asked whether the *MALT* and *MALS* families show similar sub- and/or neofunctionalization toward different carbohydrates. Phylogenetic analysis revealed three distinct *MALT* clades and five *MALS* clades (Figure 4). To investigate the specificity of individual transporters and maltases, we

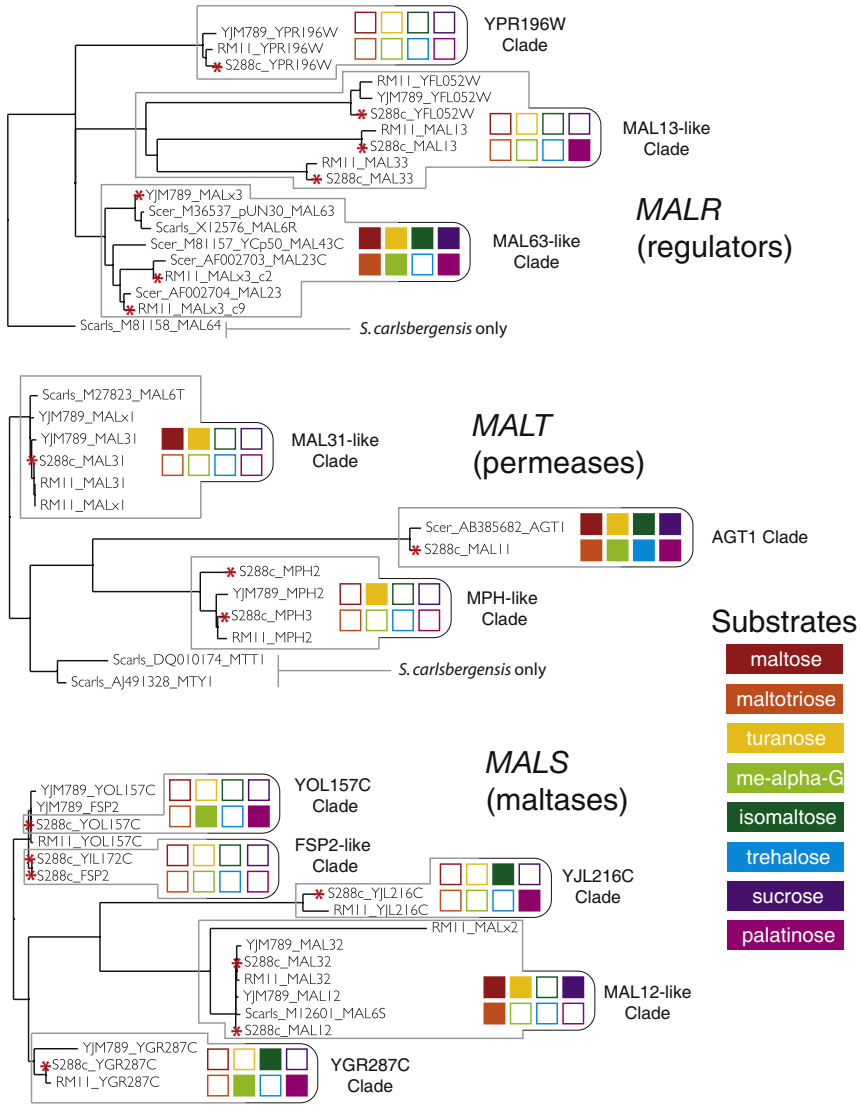

Figure 4. Phylogeny of *MAL* Genes in *S. cerevisiae*

The *MAL* gene families subdivide into tight subfamilies (clades) that correlate with specificity toward specific substrates. The phylogeny of the *MAL* regulators, transporters, and maltases, determined from multiple protein sequence alignments, is shown for alleles from *S. cerevisiae* strains S288c, YJM789, and RM11, as well as alleles deposited in GenBank from *S. cerevisiae* and *S. carlsbergensis*. Individual subfamilies are outlined with gray boxes, in which specificity is denoted by a colored barcode. A red asterisk to the left of the allele name denotes that the function of this allele was experimentally investigated. The functions of alleles that are not marked by an asterisk were not experimentally verified (in these cases, the function was only inferred from the sequence similarity with other alleles in the same clade).

The specificity of the individual families was determined as follows. Combinatorial knockouts of *MALR* alleles in S288c (Figure 5C), knockins of *MALR* alleles from RM11 and YJM789 into S288c (Figures 5A and 5B), and combinatorial knockouts of *MALR* alleles in RM11 and YJM789 (Figure 5A) were used to determine *MALR* allele specificity. Overexpression of *MALT* alleles in S288c (Figures 6A and 6B) and knockouts of *MALT* alleles in S288c (Figure 5F) were used to determine *MALT* allele specificity. Combinatorial knockouts of *MALS* alleles in S288c (Figure 5D and 5E), overexpression of *MALS* alleles in S288c (Figure 6A and 6B), and purified enzyme assays of MalS proteins (Figure 6A and 6B) were used to determine *MALS* allele specificity. For more detailed information about assays, see Experimental Procedures. Activity of a subfamily is summarized for maltose (red), maltotriose (orange), turanose (yellow), methyl-α-glucoside (lime green), isomaltose (green), trehalose (light blue), sucrose (purple), and palatinose (magenta). Activity toward a specific substrate is indicated by a solid-colored square, whereas lack of activity for a specific substrate is depicted by white boxes with colored outlines. See Figure S4 for more information.

created a yeast strain without active *MALR* genes (so that all *MALS* and *MALT* genes were silent). Using this strain, we constructed yeast mutants that constitutively express different combinations of one *MALT* and one *MALS* gene each and tested their growth on a series of carbohydrates (Figure S6). Certain combinations of *MALS* and *MALT* pairs allowed growth on specific sugars. For example, expressing *MAL11* (*MALT* family member) in combination with *MAL12* or *MAL32* (*MALS* family members) allowed S288c to grow on maltotriose, whereas expressing *MAL11* in combination with *YOL157C* or *YGR287C* (*MALS* family members) allowed growth on methyl-α-glucoside (Figure 6A). Together, the tests indicate that the different *MALT* and *MALS* subgroups allow import and hydrolysis of specific α-glucosides. Some clades encoded proteins with broad substrate specificity (e.g., *MAL11* member of the *MALT* family), whereas others were more specific (e.g., *YOL157C* member of the *MALS* family) (Figure 4; Figures 5D–5F; Figure 6).

To further confirm the substrate specificity of the *MALS* family members, we purified all seven maltase proteins (Mal12p, Mal32p, Fsp2p, Yil172cp, Yol157cp, Yjl216cp, and Ygr287cp) from *S. cerevisiae* S288c and measured their ability to hydrolyze different α-glucosides. The results confirm that *MAL12*-like clade genes (e.g., Mal12p and Mal32p) (Figure 4) have evolved specificity for maltose, maltotriose, turanose, and sucrose (Figure 5D; Figure 6A), whereas other clades (*YOL157C*, *YJL216C*, and *YGR287C*) (Figure 4) have evolved specificity for other carbohydrates, such as palatinose, isomaltose, and methyl-α-glucoside (Figure 5E; Figure 6). These results agree perfectly with the previous assays in which genes were deleted or overexpressed.

Discussion

Our results uncover the extraordinary dynamics of subtelomeric gene families. Genes residing near the telomeres undergo frequent recombination and duplication, which may allow evolutionary adaptation and innovation. Detailed analysis of three gene families that were historically linked with maltose metabolism confirms our genome-wide in silico analysis. In some yeasts, the *MAL* genes have completely disappeared, whereas in others they show multiple recent duplication events. Moreover, the evolutionary rate at which these changes have taken place is exceptional, with wide

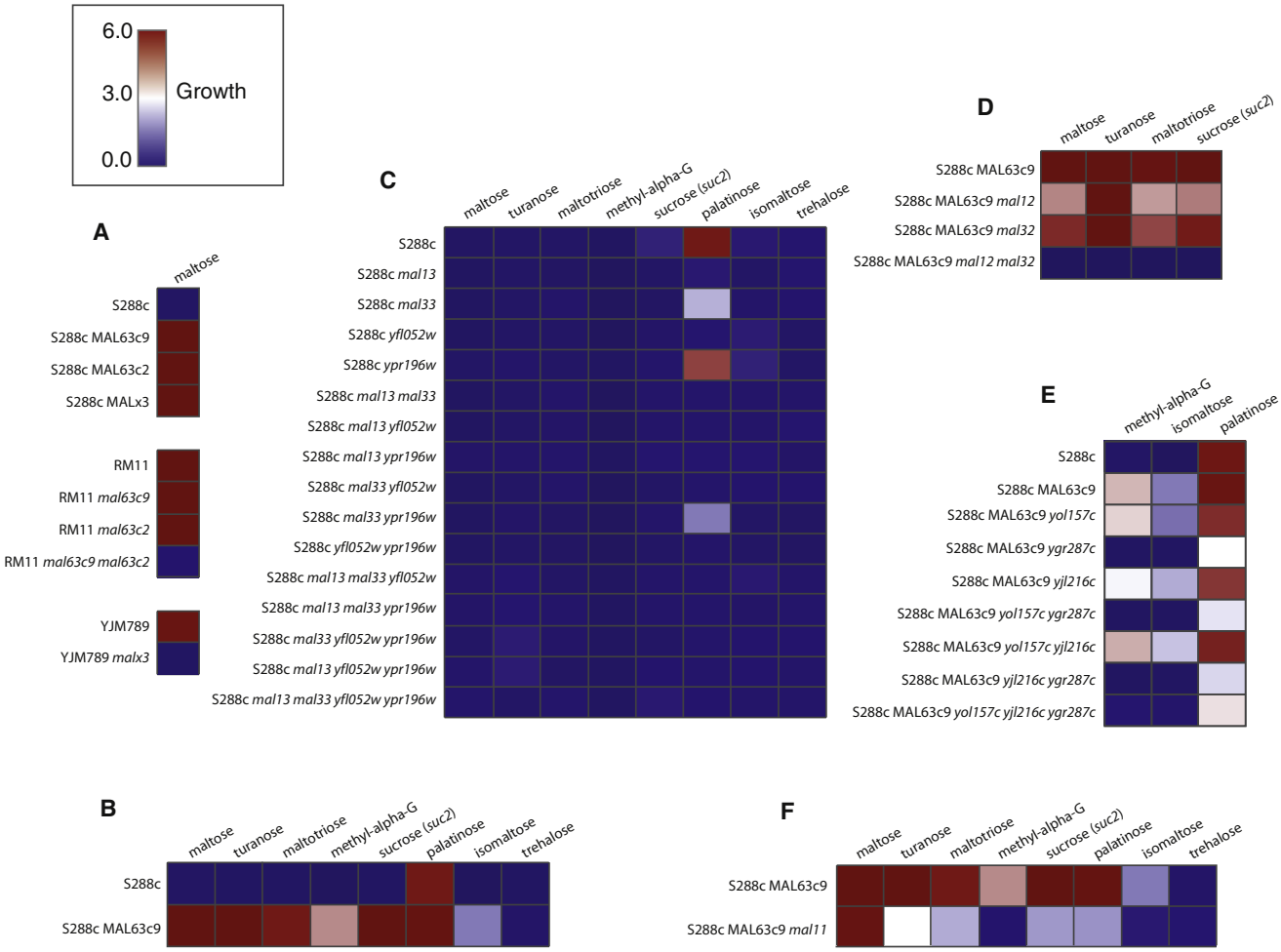

Figure 5. Growth Data for *MAL* Mutants

MAL deletion mutants confirm phenotypes and functional divergence. Growth of various *MAL* mutants is portrayed in a heat map going from no growth (dark blue, 0.0) to strong growth (dark red, 6.0).

(A) S288c wild-type strain fails to grow on maltose, whereas RM11 and YJM789 both grow on maltose. Transforming either a functional regulator from RM11 or YJM789 confers growth in S288c. Conversely, removing the functional regulator from YJM789 or both functional regulators from RM11 renders both strains unable to grow on maltose.

(B) The functional regulator from RM11, *MAL63c9* (*MAL63* found on supercontig 9 in RM11 [see Figure S5]), is not only required for growth on maltose but is also required for growth on turanose, maltotriose, methyl-α-glucoside, sucrose (*suc2* mutant), palatinose, and isomaltose.

(C) All possible combinatorial knockouts of *MAL* regulators in S288c reveal that *MAL13* and *YFL052W* are required for growth on palatinose, whereas the absence of *MAL33* reduces growth on this carbon source.

(D) Two maltases, *MALS* genes *MAL12* and *MAL32*, are required for growth on maltose, turanose, maltotriose, and sucrose (the latter tested in a *suc2* mutant), whereas the other *MALS* family members don't affect the phenotype.

(E) *YGR287C*, a *MALS* gene, is the only *MALS* family member required for growth on palatinose, isomaltose, and methyl-α-glucoside.

(F) Removal of *MAL11* permease renders strains unable to grow on most α-glucosides. See Figure S5 for more information.

differences in copy number within closely related species of the *Saccharomyces sensu stricto* group (Figure 2) and even within one species (Figure S4). Moreover, the various *MAL* loci reveal a surprisingly broad activity, with certain previously unidentified new family members showing no activity toward maltose but instead degrading several other α-glucosides. Because of the remarkable evolutionary rate of these gene families, it is difficult to predict the specificity of the ancestral enzymes, and it remains a future direction to determine the extent to which neofunctionalization and subfunctionalization have shaped their evolution.

It is interesting to hypothesize about the origins of variability of subtelomeric genes and gene families. In our analysis, we noted specific functional categories of genes that are enriched in subtelomeres. Nonsubtelomeric genes that belong to the same functional categories do not show a similar variability, suggesting that the rapid turnover of subtelomeric genes is an inherent property of these regions and not of the functional categories of genes. It is difficult to deduce whether certain rapidly evolving genes are adaptively relocated to subtelomeres or whether genes are relocated purely randomly to the telomeres, which results in rapid evolution. Although our results do not allow us to differentiate between these two scenarios, we hypothesize that both scenarios may be true. Genes may be relocated to the telomeres more or less randomly, but only those genes for which the local elevated dynamics are associated with a selective advantage will be retained in the subtelomeres.

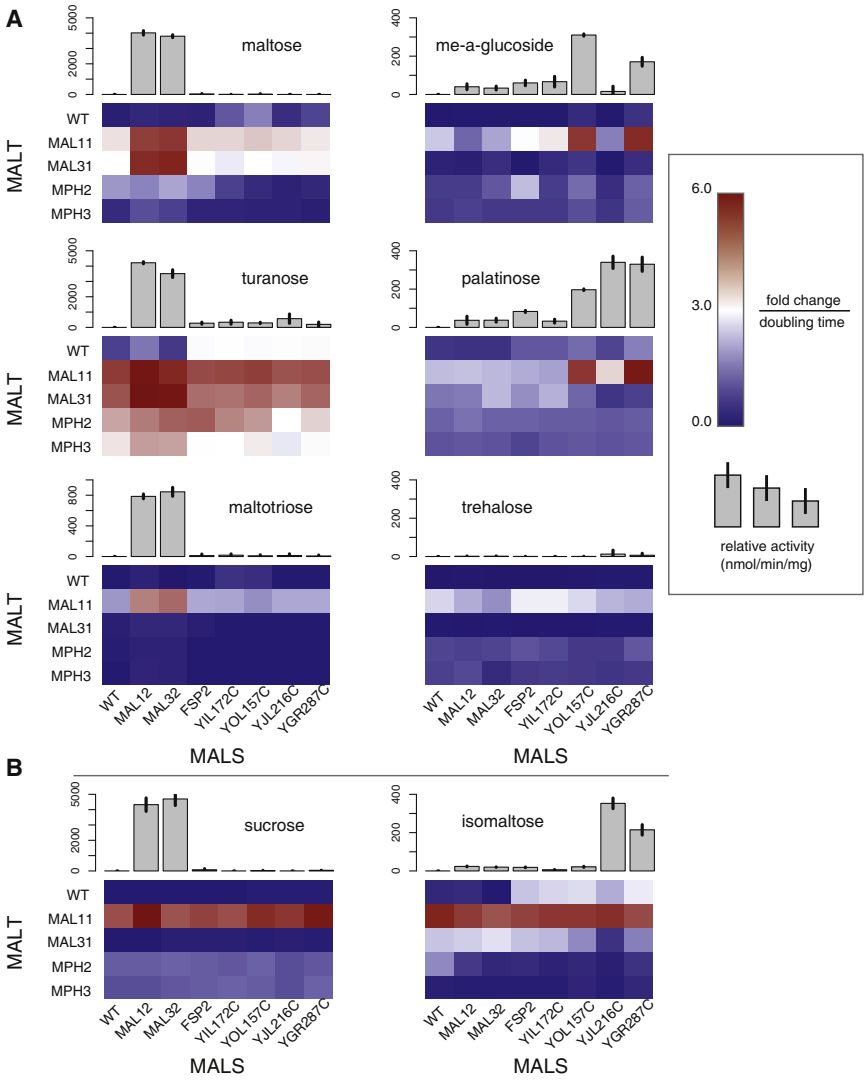

Figure 6. Growth and Enzyme Assays

Growth and enzyme assays of *MAL* overexpression mutants show the functional divergence in the *MAL* families. Growth in various α-glucosides of *S. cerevisiae* S288c diploids resulting from a cross of a *MALT* (rows) overexpression strain and a *MALS* (columns) overexpression strain is shown as a heat map. It is important to note that the upper left entry in the heat map is a wild-type S288c diploid that is a control for growth. For all of the sugars except palatinose and sucrose, the genotype of the diploids is S288c, with exception of the *MALS* and *MALT* modifications. For palatinose, S288c *mal13/mal13* diploids were used, whereas for sucrose, S288c *suc2/suc2* diploids were used. Relative activity in units of nmol/min/mg of purified *MALS* proteins is shown as gray bar graphs above the respective *MALS* column in the heat map. Error bars denote ±1 standard deviation. Individual family members are denoted beside their row or column in the heat map.

(A) Growth of the diploids in maltose, maltotriose, methyl-α-glucoside, palatinose, trehalose, and turanose correlates well with the relative activity of the purified enzymes and implicates specific combinations of *MALT* and *MALS* alleles.

(B) Growth of the diploids in sucrose and isomaltose is dependent on the *MAL11* family member of *MALT* and no specific *MALS* family members, whereas the enzyme assays indicate specific *MALS* proteins. This is most likely due to *MALR* alleles responding to the imported sugars and upregulating the pertinent *MALS* alleles once the native regulation of the *MAL11* allele of *MALT* has been bypassed. Deletion of the putative *MALS* alleles indicated by the enzyme assays confirms their specificity for sucrose (Figure 5D) and isomaltose (Figure 5E). See Figure S6 for more information.

Because most subtelomeric gene families are involved in niche-specific processes, including carbohydrate metabolism [18, 19, 21, 41, 47], stress response, and cell surface properties [22, 23, 48], it is tempting to speculate that their evolvability allows rapid adaptation to novel niches and population structures [49, 50]. In the case of the *MAL* genes, expansion of the gene families in *Saccharomyces sensu stricto* may have allowed the metabolism of carbohydrates found in plants and fruits, whereas further selection by brewers has probably led to the other observed expansions [51]. Table S5 shows a significant amplification in *S. paradoxus* of specific *MAL* alleles involved in the metabolism of sucrose, palatinose, and other sugars found in tree sap and honey, from which *S. paradoxus* is often isolated [16, 52, 53]. In yeasts that colonize mammals, such as *Candida spp*., the *MAL* genes were completely lost, presumably because these yeasts encounter enough simple, preferred sugars present in blood and the digestive tract. Similarly, expansion of subtelomeric gene families may have supported an elegant immune evasion system in pathogens, whereas the contraction of olfactory receptors in humans may explain our inability to detect certain smells [9, 26]. Interestingly, a recent study in which *S. cerevisiae* cells were evolved under sulfur-limited conditions identified frequent duplications of *SUL1*, a subtelomeric gene encoding a sulfate permease located near a *MAL* locus on the right arm of chromosome II [54].

Recent studies have noted the importance of whole-genome duplication events for evolutionary innovation [17, 42, 55, 56]. Although these duplication events are rare, our results indicate that small-scale duplication events in the subtelomeric regions may also serve an important evolutionary role. Whereas the number of subtelomeric genes is much smaller than all genes involved in whole-genome duplication events, innovation at subtelomeres is a continuous process rather than a rare event. Furthermore, subtelomeric specific epigenetic effects, including chromatin-dependent silencing, may further add to the evolutionary potential of these interesting regions, for example, by allowing swift divergence of the transcriptional regulation of the duplicated copies, a crucial but often overlooked process in evolution [57].

Experimental Procedures

Below is a summary of the experimental procedures. See the Supplemental Experimental Procedures for more details about the individual sections.

Microbial Strains, Growth Conditions, and Molecular Techniques

All yeast strains and oligonucleotides (Sigma-Genosys and IDT) used are listed in the Supplemental Experimental Procedures. Yeast cultures were

grown as described previously [58]. The sugars used in this study were purchased to their highest available purity and were filter-sterilized before adding to rich media. Standard cloning and molecular biology procedures were used [59]. Growth assays were preformed in a BioScreen C MBR system (Oy Growth Curves AB). Overnight cultures were diluted and inoculated, and after 48 hr the machine was stopped and the doubling time and fold change ($OD600_{final}/OD600_{initial}$) were calculated.

Protein Purification

Yeast strains carrying pGPD-3xHA-MalS proteins (strains KV2325–KV2331), as well as a wild-type strain (KV447), were inoculated into 1 liter of yeast extract and peptone with 2% glucose from overnight cultures to a starting OD600 of 0.1. HA-tagged proteins from the cell extracts were purified with EZview red anti-HA affinity gel (Sigma). Protein quality and quantity were determined on a NuPage Novex Bis-Tris mini gel (Invitrogen) with bovine serum albumin (BSA) standards (Fluka).

Enzyme Assays

The relative activities of the MalS proteins were determined by measuring glucose release with a GOD-PAP kit (Dialab). Reaction mixtures consisted of 3 μl of purified protein in 27 μl of a 100 mM phosphate buffer at pH 6.8 with 0.5 mg/mL BSA and 500 mM of a given sugar. The mean and standard deviation of the relative activity (nmol of product hydrolyzed/min/mg protein) were calculated from three independent reactions.

Computational Data Sets

The proteomes of the eight completely sequenced ascomycetes (*Saccharomyces cerevisiae* S288c, *Candida glabrata* CBS138, *Kluyveromyces lactis* NRRL Y-1140, *Ashbya gossypii* ATCC 10895, *Pichia stipitis* CBS 6054, *Debaryomyces hansenii* CBS767, *Yarrowia lipolytica* CLIB122, and *Schizosaccharomyces pombe* 972h-) were downloaded from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). The coordinates of the genes encoding each protein along with the respective chromosome length were used to determine relative gene coordinates, which is simply the minimal distance to nearest chromosome end. This relative coordinate was then used to determine whether a gene is subtelomeric or not based off of a cutoff (e.g., 33 kb). This was used for all of the subsequent statistics (see Supplemental Experimental Procedures).

Gene Family Statistics

Gene families were determined by using the MCL algorithm (http://www.micans.org/mcl/). Although we used an inflation parameter of 2 for most of our analyses, the MCL inflation parameter was varied (1.5–5), and our results remained unaltered (see Table S2). These gene families were then used for all of our analyses (see Supplemental Experimental Procedures).

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, six figures, and three tables and can be found with this article online at doi:10.1016/j.cub.2010.04.027.

Acknowledgments

The authors thank Verstrepen Laboratory members, G.R. Fink, E.K. O'Shea, C. Michels, M. Hahn, S. Maere, and B. Stern for their valuable advice. C.A.B. acknowledges financial support from the National Science Foundation, Harvard University Integrated Life Sciences, and a Harvard Sheldon Fellowship. Research in the laboratory of K.J.V. is supported by National Institutes of Health grant P50GM068763, Human Frontier Science Program RGY79/2007, European Research Council Young Investigator grant 241426, VIB, K.U.Leuven, the Fonds Voor Wetenschappelijk Onderzoek – Vlaanderen (FWO)-Odysseus program, and the AB InBev Baillet-Latour foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Received: February 8, 2010
Revised: April 13, 2010
Accepted: April 14, 2010
Published online: May 13, 2010

References

1. Pryde, F.E., Gorham, H.C., and Louis, E.J. (1997). Chromosome ends: All the same under their caps. Curr. Opin. Genet. Dev. *7*, 822–828.
2. Cohn, M., Liti, G., and Barton, D.B.H. (2006). Telomeres in fungi. In Comparative Genomics Using Fungi as a Model, *Volume 15*, P. Sunnerhagen and J. Piskur, eds. (Heidelberg, Germany: Springer), pp. 101–130.
3. Mefford, H.C., and Trask, B.J. (2002). The complex structure and dynamic evolution of human subtelomeres. Nat. Rev. Genet. *3*, 91–102.
4. Pryde, F.E., and Louis, E.J. (1999). Limitations of silencing at native yeast telomeres. EMBO J. *18*, 2538–2550.
5. Gottschling, D.E., Aparicio, O.M., Billington, B.L., and Zakian, V.A. (1990). Position effect at S. cerevisiae telomeres: Reversible repression of Pol II transcription. Cell *63*, 751–762.
6. Rudd, M.K., Friedman, C., Parghi, S.S., Linardopoulou, E.V., Hsu, L., and Trask, B.J. (2007). Elevated rates of sister chromatid exchange at chromosome ends. PLoS Genet. *3*, e32.
7. Linardopoulou, E.V., Williams, E.M., Fan, Y., Friedman, C., Young, J.M., and Trask, B.J. (2005). Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature *437*, 94–100.
8. Barton, A.B., Pekosz, M.R., Kurvathi, R.S., and Kaback, D.B. (2008). Meiotic recombination at the ends of chromosomes in Saccharomyces cerevisiae. Genetics *179*, 1221–1235.
9. Mefford, H.C., Linardopoulou, E., Coil, D., van den Engh, G., and Trask, B.J. (2001). Comparative sequencing of a multicopy subtelomeric region containing olfactory receptor genes reveals multiple interactions between non-homologous chromosomes. Hum. Mol. Genet. *10*, 2363–2372.
10. Anderson, J.A., Song, Y.S., and Langley, C.H. (2008). Molecular population genetics of Drosophila subtelomeric DNA. Genetics *178*, 477–487.
11. Kern, A.D., and Begun, D.J. (2008). Recurrent deletion and gene presence/absence polymorphism: Telomere dynamics dominate evolution at the tip of 3L in Drosophila melanogaster and D. simulans. Genetics *179*, 1021–1027.
12. Britten, R.J. (1998). Precise sequence complementarity between yeast chromosome ends and two classes of just-subtelomeric sequences. Proc. Natl. Acad. Sci. USA *95*, 5906–5912.
13. Eichler, E.E. (2001). Segmental duplications: What's missing, misassigned, and misassembled—and should we care? Genome Res. *11*, 653–656.
14. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241–254.
15. Hahn, M.W., De Bie, T., Stajich, J.E., Nguyen, C., and Cristianini, N. (2005). Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Res. *15*, 1153–1160.
16. Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., et al. (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337–341.
17. Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. Nature *449*, 54–61.
18. Carlson, M., Celenza, J.L., and Eng, F.J. (1985). Evolution of the dispersed SUC gene family of Saccharomyces by rearrangements of chromosome telomeres. Mol. Cell. Biol. *5*, 2894–2902.
19. Denayrolles, M., de Villechenon, E.P., Lonvaud-Funel, A., and Aigle, M. (1997). Incidence of SUC-RTM telomeric repeated genes in brewing and wild wine strains of Saccharomyces. Curr. Genet. *31*, 457–461.
20. Michels, C.A., and Needleman, R.B. (1984). The dispersed, repeated family of MAL loci in Saccharomyces spp. J. Bacteriol. *157*, 949–952.
21. Naumov, G., Turakainen, H., Naumova, E., Aho, S., and Korhola, M. (1990). A new family of polymorphic genes in Saccharomyces cerevisiae: Alpha-galactosidase genes MEL1-MEL7. Mol. Gen. Genet. *224*, 119–128.
22. Teunissen, A.W., and Steensma, H.Y. (1995). Review: The dominant flocculation genes of Saccharomyces cerevisiae constitute a new subtelomeric gene family. Yeast *11*, 1001–1013.
23. Verstrepen, K.J., and Klis, F.M. (2006). Flocculation, adhesion and biofilm formation in yeasts. Mol. Microbiol. *60*, 5–15.
24. Verstrepen, K.J., and Fink, G.R. (2009). Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. Annu. Rev. Genet. *43*, 1–24.

25. Barry, J.D., Ginger, M.L., Burton, P., and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? Int. J. Parasitol. *33*, 29–45.
26. Hasin, Y., Olender, T., Khen, M., Gonzaga-Jauregui, C., Kim, P.M., Urban, A.E., Snyder, M., Gerstein, M.B., Lancet, D., and Korbel, J.O. (2008). High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. PLoS Genet. *4*, e1000249.
27. Linardopoulou, E.V., Parghi, S.S., Friedman, C., Osborn, G.E., Parkhurst, S.M., and Trask, B.J. (2007). Human subtelomeric WASH genes encode a new subclass of the WASP family. PLoS Genet. *3*, e237.
28. Knight, S.J.L., and Flint, J. (2000). Perfect endings: A review of subtelomeric probes and their use in clinical diagnosis. J. Med. Genet. *37*, 401–409.
29. Bailey, J.A., and Eichler, E.E. (2006). Primate segmental duplications: Crucibles of evolution, diversity and disease. Nat. Rev. Genet. *7*, 552–564.
30. Fairhead, C., and Dujon, B. (2006). Structure of Kluyveromyces lactis subtelomeres: Duplications and gene content. FEMS Yeast Res *6*, 428–441.
31. Gordon, J.L., Byrne, K.P., and Wolfe, K.H. (2009). Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern Saccharomyces cerevisiae genome. PLoS Genet. *5*, e1000485.
32. Louis, E.J. (1995). The chromosome ends of Saccharomyces cerevisiae. Yeast *11*, 1553–1573.
33. Winzeler, E.A., Castillo-Davis, C.I., Oshiro, G., Liang, D., Richards, D.R., Zhou, Y., and Hartl, D.L. (2003). Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics *163*, 79–89.
34. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al; The Gene Ontology Consortium (2000). Gene ontology: Tool for the unification of biology. Nat. Genet. *25*, 25–29.
35. Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.
36. Van Dongen, S. (2008). Graph clustering via a discrete uncoupling process. Siam J. Matrix Anal. Appl. *30*, 121–141.
37. Schober, H., Kalck, V., Vega-Palas, M.A., Van Houwe, G., Sage, D., Unser, M., Gartenberg, M.R., and Gasser, S.M. (2008). Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. Genome Res. *18*, 261–271.
38. Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics *151*, 1531–1545.
39. Ohno, S. (1970). Evolution by Gene Duplication (New York: Springer).
40. Tirosh, I., Weinberger, A., Carmi, M., and Barkai, N. (2006). A genetic signature of interspecies variations in gene expression. Nat. Genet. *38*, 830–834.
41. Charron, M.J., Read, E., Haut, S.R., and Michels, C.A. (1989). Molecular evolution of the telomere-associated MAL loci of Saccharomyces. Genetics *122*, 307–316.
42. Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature *428*, 617–624.
43. Han, E.K., Cotty, F., Sottas, C., Jiang, H., and Michels, C.A. (1995). Characterization of AGT1 encoding a general α-glucoside transporter from Saccharomyces. Mol. Microbiol. *17*, 1093–1107.
44. Cheng, Q., and Michels, C.A. (1991). MAL11 and MAL61 encode the inducible high-affinity maltose transporter of Saccharomyces cerevisiae. J. Bacteriol. *173*, 1817–1820.
45. Jespersen, L., Cesar, L.B., Meaden, P.G., and Jakobsen, M. (1999). Multiple alpha-glucoside transporter genes in brewer's yeast. Appl. Environ. Microbiol. *65*, 450–456.
46. Michels, C.A., and Needleman, R.B. (1983). A genetic and physical analysis of the MAL1 and MAL3 standard strains of Saccharomyces cerevisiae. Mol. Gen. Genet. *191*, 225–230.
47. Ness, F., and Aigle, M. (1995). RTM1: A member of a new family of telomeric repeated genes in yeast. Genetics *140*, 945–956.
48. Rachidi, N., Martinez, M.J., Barre, P., and Blondin, B. (2000). Saccharomyces cerevisiae PAU genes are induced by anaerobiosis. Mol. Microbiol. *35*, 1421–1430.
49. Gore, J., Youk, H., and van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. Nature *459*, 253–256.
50. Rando, O.J., and Verstrepen, K.J. (2007). Timescales of genetic and epigenetic inheritance. Cell *128*, 655–668.
51. Dunn, B., and Sherlock, G. (2008). Reconstruction of the genome origins and evolution of the hybrid lager yeast Saccharomyces pastorianus. Genome Res. *18*, 1610–1623.
52. de la Fuente, E., Sanz, M.L., Martinez-Castro, I., Sanz, J., and Ruiz-Matute, A.I. (2007). Volatile and carbohydrate composition of rare uniflo-ral honeys from Spain. Food Chem. *105*, 84–93.
53. Heizmann, U., Kreuzwieser, J., Schnitzler, J.-P., Brueggemann, N., and Rennenberg, H. (2001). Assimilate transport in the xylem sap of pedunculate oak (Quercus robur) saplings. Plant Biol. *3*, 132–138.
54. Gresham, D., Desai, M.M., Tucker, C.M., Jenq, H.T., Pai, D.A., Ward, A., DeSevo, C.G., Botstein, D., and Dunham, M.J. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. *4*, e1000303.
55. Fawcett, J.A., Maere, S., and Van de Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc. Natl. Acad. Sci. USA *106*, 5737–5742.
56. Scannell, D.R., and Wolfe, K.H. (2008). A burst of protein sequence evolution and a prolonged period of asymmetric evolution follow gene duplication in yeast. Genome Res. *18*, 137–147.
57. Tirosh, I., Barkai, N., and Verstrepen, K.J. (2009). Promoter architecture and the evolvability of gene expression. J. Biol. *8*, 95.
58. Abelson, J., Simon, M., Guthrie, C., and Fink, G. (2004). Guide to Yeast Genetics and Molecular Biology (New York: Academic Press), pp. 933.
59. Fritsch, E., Maniatis, T., Sambrook, J., and Russell, D. (2001). Molecular Cloning: A Laboratory Manual (Woodbury, NY: Cold Spring Harbor Laboratory Press), pp. 999.